

Association Rule Mining: An Application Perspective

Kamran Shaukat, Sana Zaheer, Iqra Nawaz

Department of Information Technology, University of the Punjab, Jhelum Campus, Pakistan.

Email: *Kamran, bcs.f12.30, bcs.f12.03@pujc.edu.pk*

Abstract- Data mining is an emerging field and it is a method to find out interesting patterns and knowledge from the large amount of sales data in the transactional and interpersonal database .The most important and extremely studied function of data mining is Association Rule Mining (ARM). We have proposed a detailed review of ARM applications and presented the various areas in which association rules are applied for extracting interesting patterns and relationships among set of items in data repositories. The main purpose of this function is to find frequent patterns, associations and relationship between various database using different Algorithms. ARM is used to improve decisions making in the applications such as Market basket analysis, Disease diagnosis, building Intelligent transportation system(ITS), Web log data, Computer aided diagnosis system of Breast cancer, fraud detection in web, CRM of credit card business by using different Algorithms such as Apriori, FP-Growth, Equivalence class transformation(ECLAT), Direct hashing and pruning, Tree Projection.

Keywords: *Frequent Pattern mining; Data Mining; Association rule mining; Apriori Algorithm.*

I. INTRODUCTION

Data mining (DM) is a process of discovering fascinating designs, new instructions and information from large amount of sales facts in transactional and interpersonal catalogs. It is an emerging technology which involves the artificial intelligence, learning of machine, information retrieval and high performance computing. It provides important and valuable information/knowledge to decision makers, resulting in incalculable economic benefits. The functions of data mining are cluster analysis, outlier analysis, classification and prediction, association and correlations analysis. Most of the work in ARM is done by Agrawal. It is the most important and deeply considered functions of mining of data. It not only provides a well-organized method of discovering the patterns and recognition of the model but also verifies the rule that exist, which ultimately helps in providing new rules. Association rules provide the effective scientific base for decision making. Rule of associations have been used in many applications to find frequent patterns in data .One of the key domains which use association rule is business field where it helps in a very effective and efficient decision making and marketing. Other field areas where association rule mining can be applied are market basket analysis, medical diagnosis, census data, fraud detection in web and DNA data analysis etc. For example in the field of electrical power engineering, the methods of data mining have been used for the condition monitoring of high voltage equipment. In medical field ARM is used to find frequently occur diseases in particular area and to diagnose different diseases. It is also used to attain information about the navigational activities of users in web Log data. Recently it is discovered that there are various algorithms for finding the association rules. For frequent pattern mining different frameworks have been defined. One of the best and most commonly used

algorithm is Apriori. Other procedures are Equivalence Class Transformation (ÉCLAT), Tree Projection, FP-growth, direct hashing and pruning algorithm

We summarize the paper in this as follow: Section 1 describes the tasks of ARM. Section 2 describes the literature review and in Section 3 the applications of ARM are described in detail.

II. LITERATURE REVIEW

Data excavating is stated as information/knowledge detection in databases (KDD). It is a procedure that abstracts hidden and beforehand indefinite valuable evidence. It is the method of analyzing large quantity of facts from dissimilar perspective, after that succinct it into valuable info. The info can transform into knowledge about the past patterns which helps in understanding expected future trends. The amount of data generate by a variety of disciplines is increasing speedily, resulting in difficulties in finding useful knowledge. In order to solve this problem Data mining (DM) methods are used. Data mining (DM) techniques are deployed to score huge database to find the valued patterns that might remain unrevealed, if these methods are not used [1]. It is the method of determining fascinating patterns, new rules and all the information from a huge amount of facts in the transactional and interpersonal databases [2]. Data excavating knowledge has obvious benefits in huge data dispensation, possible evidence to find the rulebooks of association discovery. Functions of DM are association rule mining (ARM), cluster analysis, outlier analysis, classification and prediction, generalization and characterization. Connotation rule(AR) is the significant and the most dynamic acquaintance models of data excavating. Most of the work in ARM is done by Agrawal. Detail review of literature has lead the researchers to the fact that ARM is one of the most significant data mining field which is used by researchers in numerous areas of retailing. Liao's offered cross sales recommendations for the electronic catalogue designs, by using association rules based on data mining method [3]. Association rule(AR) denotes relates relationship of a group in database. It emphasizes on the connection of dissimilar attribute area, which can encounter positive necessities. It can be recycled to recover the conclusion making process in variety of solicitations as MBA, medical analysis of sicknesses, protein categorizations, logistic deterioration and deceit recognition in net log data and CRM of credit card business, building intelligent transportation system(ITS) and intelligent traffic light control.

There are various algorithms for finding Association Rule (AR) such as Equivalence class transformation (Éclat), Tree Projection, FP-growth, Apriori, Rapid association rule mining (RARM), Associated Sensor Pattern Mining (ASPMS), Direct hashing and Pruning procedure, mine algorithm, Partitioning algorithm, Dynamic item set counting algorithm, Moment algorithm.

- The basic join based procedure is the Apriori method. In Apriori procedure, breadth first and pruning property has been used. It is easy to implement and use large item set property. Too many passes over database require large database.
- In Rapid Association Rule Mining (RARM), depth first or sotrielt to generate 1-itemset & 2-itemset technique used. It has No candidate generation and speed up process for generating item set 1 & 2. It is difficult to use in interactive mining and incremental mining. As compare to apriori there is no candidate generation in FP growth algorithm. It uses divide and conquer technique.
- H-mine algorithm uses hyperlink pointers for storing the partitioned projected database in main memory. It is suitable for sparse and dense database. Dynamic Item set counting algorithm is based upon dynamic insertion of candidate. It is suitable for medium and low database.
- Direct hashing and pruning algorithm use hashing technique for finding frequent itemset. It is suitable for medium database.

ARM is broadly used in many applications for finding pattern in data mining.

- ARM is used in market basket exploration to determine the relationships among the items that are purchased together by the customer in transaction database. The item's place in the store can create important differences in its sales. Therefore the information of the items which are sold will show which items should put side by side in store shelves. Data mining methods are used to find the collection of items, which are purchased together. In market basket analysis, apriori algorithm can be used because it is candidate generation algorithm.
- In medical field, using the improved apriori algorithm we can find the recurrent element sets in a catalogue of the medicinal verdict, and makes the robust AR in edict to find out connotation association or configurations among the huge data item sets. It shows that the amended apriori procedure can take out the connotation rule prototypes about belongings and nature of the sickness from the medicinal catalogues, which can help specialists in medical analysis. In the meadow of medicinal the data excavating applications are DNA sequence analysis, neural network in clinical analysis, association analysis in medicine, disease diagnosis.
- The application of recurrent item mining is to discover recurrent configurations in Net log facts. The three configurations to be scrutinized are recurrent item sets, categorizations and tree arrangements. For each problem a procedure was established to discover the configurations powerfully. The recurrent item sets (frequent page sets) are concealed by using the Itemset Code procedures-M Tree procedure uses state-run machines to discover categorizations, and P-D Tree algorithm uses Pushdown automatons to determine the support of tree configurations in tree catalogue.
- One of the applications of frequent item mining is building intelligent transportation system. It can be built using FP-growth and unid_FP-Max. FP-growth has higher operating efficiency than the Unid_FP-Max in building intelligent transportation system.
- Another application is configuration detection method to trade the deceit discovery. A few automatic video schemes have been established to display the treasurer paths and perceive non-compliant accomplishments. These schemes use facts from security cinematic cameras and (TLog) that are verified at POS. A pattern discovery method is to identify fake events at POS. It establish mining time ordered text creeks, representing retail communications, shaped from the arrangement of visually perceived the point of sale associated actions named the primitives and barcodes since the TLog facts is used.

Many other methods have been used in other applications of the frequent item mining.

III. APPLICATIONS

A. *Market Basket Analysis (MBA)*

One of the most common applications of ARM is market basket analysis (MBA) that discovers the relations among the items obtained by customers in the database. The improvement in the information technology allows all the retailers to obtain the daily transaction data at a very low cost. Thus, the large amount of useful data to support the retail management can be extracted from large transactional databases. Data mining (DM) is used to obtain valuable information from large databases [4]. The aim of ARM analysis is to describe the most interesting patterns in an efficient manner [5]. ARM analysis (also known as the market basket analysis (MBA)) is method of determining customer obtained patterns by mining association from retailer transactional database [6]. Now a day's every product comes with the bar code. This data is rapidly documented by the business world as having the huge possible value in marketing. In detailed, commercial organizations are interested in "association rules" that identify the patterns of purchases, such that the occurrence of one item in a basket will indicate the presence of one or more additional items. This "market basket analysis" result can then be used to recommend the combinations of the products for special promotions or sales, devise a more actual store layout, and give vision into brand loyalty and co-branding. It will also lead the managers towards efficient and real strategic decision making. Data mining(DM) methods are also used to find the collection of products, which are purchased together. It helps to choose which products should put

side by side in the store shelves which may lead to important increase in sales. The problem of ARM can be decayed into the succeeding two stages [7]

- a) Determine large item sets, i.e. item sets that have operation support above a predetermined min support. Item sets which have the min support are called the recurrent item sets.
- b) Large item sets to produce association rules for database.

Algorithm used in market basket analysis (MBA) is apriori algorithm because it is a candidate generation algorithm. It is founded on information that this algorithm uses the preceding knowledge of the regular item set possessions. Apriori procedure pays to an iterative tactic that is recognized as a level wise search in which k-item sets are used to discover (k+1) itemsets. Based on this possession, if a set cannot pass the minimum verge than all of its super sets will also fail the test as well. Thus, if an item set is not a recurrent item set, then item set will not use to create large item set [8]. Apriori procedure is the most recurrently used algorithm among the association rules algorithms that were used at the analysis phase. The problems occur in apriori algorithm are that it scans the databases again and again to check the recurrent item sets and it also generate infrequent itemsets. Strong associations have been observed among the purchased item sets group with regard to the purchase behaviour of the customers of the retail store. The customer's shopping information analyzed by using the association rules mining with the apriori algorithm. As a result of the analysis, strong and useful association rules were determined between the product groups with regard to understanding what kind of purchase behaviour customer's exhibit within a certain shopping visit from both in-category and from different product categories for the specialty store.

B. Building Intelligent Transportation System

Intelligent transportation system (ITS) is innovative information technology, beam technology, switch technology and processor technology integration that is applied to the entire transportation system. Intelligent transportation system is build within versatile role, accurate, actual time and well-organized integrated transportation controlling system [9]. Advanced traffic information service system(ITS) is assembled on informational network and it is manufactured on the basis of traffic participants on the road, parking and weather center sensors, car, transfer station, and the transmission equipment to transport information center throughout the traffic information. The system gets all the information by processing real time road traffic information to traffic participants, parking information and all other information related to travel, based on the information to determine the selection of routes. When the car is equipped with navigation system then this system can help the driver to select the exercise route automatically. Information transportation system (ITS) can efficiently decrease the traffic burden and environmental contamination. Its basic purpose is to confirm the traffic security and to recover transport effectiveness. ITS areas are advanced traveler information system, progressive public transportation system, emergency organization system, innovative vehicle control system, electronic toll group system etc. The application of ITS should have following conditions: (1) Collected road information and the traffic information should accurate, complete and real time; (2) Exchange of information between traffic management ,road management facilities should be real time and efficient ; (3) the traffic management centers , toll management centers should have self-learning computer systems. Code division multiple access(CDMA) is an intelligent traffic control system. It is divided into three parts namely traffic management centers, wireless routers, including signal control, vehicle detection, electronic police. CDMA wireless router is used to download and upload the information and the site equipment is used to complete the live traffic information collection and also controls the traffic lights.

ARM is function of DM that is used to discover the exciting connotation from the huge amount of data. In ARM the drawback of mining frequent itemsets is to consider too many candidate item sets [10]. Instead of finding recurrent itemsets only the maximal recurrent item sets that can be found. Many data mining applications only

discover the max frequent item sets without discovering all frequent item sets. Recently a lot of research has been conducted which only explain the problem of mining maxrecurrent itemsets and have not focused on mining recurrent patterns.

Many Algorithms have been suggested in Data mining to discover the frequent itemsets. Aproori Algorithm is most frequently used algorithm that uses the Breadth first and pruning technique to find frequent itemset by candidate generation. Its drawback is that it requires too many passes over database and too many candidate generations. The problem of ARM in the Transaction Database is discovering a handler given minsupport and min confidence connotation rules [11].Support characterization describes the frequency of the rules and confidence characterization describes the strength of the rules. If the degree of confidence is higher than the rule will be reliable. Association rules are very important to determine the min-support and min-confidence verge. If the threshold will be too large then the corresponding association rule will be lost and if the threshold will be too small then the corresponding association rules will not exist objectively and it will generate a lot of useless rules.

The method developed procedure based on FP-growth checks provisional pattern base to build its restricted pattern tree and implementation of mining in this tree recursively. In FP-growth there is no need to generate the candidates, the mining operation is just to calculate itemset support count and adjust the prefix tree, and this cost is usually less than the Aproori candidate itemset generation and pattern matching operations. In FP growth divide and conquer procedure is used and it provides the frequent itemset database that is flattened into FP-tree then this compacted catalog into set of restricted database every related by a common element.

Physical model that describes the real objects and their behavior, commonly used in traditional control theory, algebraic equations, differential equation etc. is introduced to compensate the drawbacks of logical model. The advantage of physical model is much simpler than the conventional mathematical model and it is used to represent intelligent transport system with a simple method. Advanced traffic management system is used to monitor and control the management of road traffic and also used to provide the communication links between driver, road and vehicle. First vehicle Auxiliary safe driving system has different sections such as vehicle sensors, microwave radar, laser radar and other forms of sensors that are used for the determination of the distance with the vehicle in front, the board computer for processing, to warn the driver in case of emergency or to force the vehicle break. New intelligent transportation system is used to improve the traffic conditions, to reduce travel time, to reduce fuel consumption and improve traffic safety and to protect the urban environment.

In this application researcher built the intelligent transportation system based on association rule mining(ARM) and FP-growth algorithm and Unid_FP-Max and the comparative analysis of the result displays that the FP-growth has advanced operating effectiveness than the UNID_FP-Max.

C. Web Log Data

The enormous use of the internet has completed the automatic knowledge extraction from web log files. Information suppliers are interested in all those techniques that could study web user's information needs and the preferences. It can be used to improve the efficiency of the websites by acclimatizing the informational construction of the places to the user's behaviour [12]. However, it is hard to discover the suitable tools for examining the rare web log data to save the important and valuable information. Currently, there are numerous general web log study tools but most of them are unpopular by the users and they measured them very slow, strict, expensive and difficult to maintain or a very partial in the results. Recently, the start of data mining techniques [13] for discovering the procedure patterns from Web data (also known as web log mining or web usage mining) specifies that these techniques can be obtainable other to traditional decision-making tackles. Common pattern mining is a very deeply researched area in the area of data mining and one of them is to use the common pattern discovery approaches in Web log data.

Discovering unseen information from the Web log data is called the web procedure mining. The goal of discovering the frequent patterns in the Web log data is to get the information about the directional behavior of the users. It can be used for marketing purposes and for making the active user profiles. There are three pattern mining

methods that are examined from the Web usage excavating. Web content mining [14, 15] is the mission of learning valuable information obtainable online. Web mining can be considered into three classes created on which part of the Web is to be mined [16, 17, and 18]. The categories are (i) Web content mining (ii) Web structure mining (iii) Web usage mining. There are dissimilar types of web content which can offer valuable information to the users, for example the software data, organized, semi-structured and unstructured data. The main goal of the web content excavating is to deliver an well-organized mechanism to relief the users to discover the information they pursue.

Dissimilar patterns in web log removal are page sets, page sequences and page graphs. Three patterns to be searched are the common item sets, sequences and tree patterns. The algorithms were established in demand to learn the common patterns efficiently.

The frequent item sets (frequent page sets) are exposed by using the ItemsetCode procedure reachable in [19]. The main benefit of the Itemset Code procedure is that it determines the small common itemsets in a rapid way, so that the mission of determining large itemsets improved as well. The procedure that determines the common page orders is called SM-Tree procedure [20] and the procedure that determines the tree-like designs is called PD Tree procedure [21]. Both of these procedures deed the advantage of using the theory of automata methods for determining all the common patterns. The SM-Tree procedure uses the state machines for shaping the orders, and the PD-Tree procedure uses the pushdown automatons for resolving the support of the tree patterns in a tree.

The input arrangement of the SM-Tree procedures covers the rows of transactions, where each row comprises a order, where the item sets are divided by a -1 value. The PD-Tree procedure plans a new procedure for defining whether the tree is confined by another tree. It can be done by using the pushdown automaton (PDA). In order to deliver an input to the automaton, the tree can be embodied as a string. To handle the huge number of candidates competently the joint operation between the automatons were suggested, and the ensuing new structure is called PD-Tree. The new erection makes it possible to determine the support of each candidate at the same time by meting out the items of a transaction exactly once. The advantage of the PD-Tree is that it uses only the one stack to achieve the removal process. It is good PD-Tree used instead of using numerous pushdown automatons(PDA). The input arrangement of the procedure also holds the rows of dealings where each deal contains a tree. A tree is signified with strings.

Basically, the item set, PD-Tree and SM-Tree procedures converses the well-organized mining measures and the common pattern removal in web log data, which discovers techniques at numerous estimate and at concept levels. There are several advantages of the tree based methods over others, as it concepts a highly dense P-tree which is usually considerably smaller than the unique database and thus it saves the rate of following removal process. This technique also compacts with the belongings where multiple concept hierarchies exist.

The difficulty of determining the hidden info from the huge amount of web log data is composed by web servers. It presents the procedure of web log drawing out, and it shows how the common pattern discovery responsibilities can be useful on the web log data in command to get the useful information about the user's steering behavior.

D. Identification Of Frequent Disease

Data mining is a procedure of describing or extracting interesting information, information or patterns from data in large database. There are many data mining techniques that are used for extraction of knowledge or patterns from data in many fields. In medical fields data mining methods are used to take medical choices in curative diseases. In health-care industries data about healthcare is collected and if that collected data is not mined that will not expose any information for real decision making. The extracted data can be used by the health-care administrators to recover the excellence of service. In the application the investigators had identified frequently occurring diseases in a specific geographical area at a specific time by using association rule(AR) based aproiri Procedure.

Association rule is the most dynamic and important knowledge models of data extraction [22]. The main focus of association rule is on the association of the quality domain that can see the certain necessities. The patterns expose the combination of the events that occur at the same time. It provides well-organized method of pattern discovery and model credit.

In diseases diagnosis association rules(AR) can be used that will be helpful for supporting physicians to treat patients. Practically disease analysis is not a easy progression as it may contain faulty diagnosis test and the presence of the noise in training examples. In this application apriori algorithm has been used to recognize recurrently occurs diseases in fastidious geographic area. Aproiri is the most standard algorithm for excavating common item sets. The apriori procedure uses previous facts of itemset belongings.

In this research the researcher serene the data from hospital information system that contains the patient's information such as patient's name, age, gender, disease etc. After collecting the entire patient's data from the hospital information system, this research can find the frequently occurred diseases using ARM. Hospital information system in medical field is used to receive the record of its patients and diseases and it is not an easy task to identify the frequent diseases without association rules as it contains a large data set. People from different locations move toward different hospitals; their record is maintained in the hospitals where they get treated. It is not easy to scrutinize all the record to find the frequently occurred diseases.

By using apriori it is easy to discover frequently occurs diseases in particular geographical area. In this application researcher composed data and analyzed that 1216 patients affected by 29 different diseases in year 2012 and by using apriori that is association based technique, analyzed that 4 different diseases occur frequently at particular geographical area in year 2012.

E. CAD (Computer Aided Diagnostic System) of Breast Cancer

With the extensive applications of computer and knowledge, the amount of data produced by many disciplines has enlarged quickly. In order to mine valuable knowledge from that data, DM methods are recycled. By using data removal techniques, hopeful results have been obtained in diseases diagnosis, organ transplantation, the treatment, image examination, drug growth, scientific study etc. In the Southern California University, Spine hospital applied data removal techniques to oncology, liver pathology, gynecology, diagnosed cases of thyroid, urology, neuropsychology and other medical fields [23]. The presentation difficulties of data excavating in Deoxyribonucleic Acid (DNA) data examination field such as compare the DNA sequence similarity, semantic integration of distributed data, genes that cause the diseases etc. was elaborated by Jiawei Han. The prediction of assembly of proteins conferring to amino acid order info was proposed by Muggleton. According to analysis of ECG, a choice tree to reset the neural system was planned by Miroslav Kubat. For the examination of Protein drug growth Vyris proposed neural network technology. Chen xuefeng used the data extraction technologies to figure a hematologic distortions analysis system that has influential analysis and removal abilities on the data [24]. In health field data digging applications has its own benefits because the data calm is overall, true, reliable and not achieved by other issues. A medical database is a enormous database. For instance a patient presenting symptoms to some disease, the laboratory tests and the treatment related to that disease may be precisely the same.

In earlier data excavating the vast number of imperfect and noisy info must be cleaned to safeguard the data reliability. The medical verdict expert systems are recycled to turn experts' knowledge into the directives. As we work the patient indications into the scheme, it will be capable to make the judgment about disease quickly and it reduces the error of doctor's slanted judgment. Data extraction techniques are also used to find the esteemed diagnostic instructions by dispensation a vast data of patient database. Data excavating techniques are used to make medical decisions to cure diseases.

Artificial neural network is an exact model to observe the working code of human mind neurons. It is the most progressive artificial intellect (AI) technology currently. It has higher performance than other conventional and mathematical models because of its self-learning and recollection ability.

In some urbanized countries CAD got dissolute expansion in the area of medical imaging [25]. The finishing analysis of CAD is dogged by the doctors and checking the production of the processor, which creates analysis outcome more exact. In health imaging the processor gives the productivity by examining the associated copy facts which helps to recover the diagnostic exactness on image and disease. CAD can recover the indicative accurateness because the radiologist's diagnostic is slanted and in doctor diagnosis the subtle changes can easily missed. Doctors have different changes on duplicate understanding and processors have a countless benefit to remove these mistakes and lacks. The aim of CAD is to make the diagnosis path more objective, calmer and efficient. It is also used to train new inexpert surgeons.

In this application researcher advanced a CAD of breast disease and advanced a consistent rule base by finding the data driven rules by database, analyzing these instructions by health experts, determining the rules opposed with knowledge. The law that is found using misleading the illness is to be excluded also should spread the training facts [26]. Currently many researchers have focused on DNA data analysis. The outcomes of DNA study exposed numerous diseases and incapacities genes, approaches for the diseases analysis and methods for the prevention and treatment. The sequence that occur frequently in disease sample exceed its in health sample are considered as disease genes and the order that occur frequently in health example beat it's in disease sample are careful as diseases fight genes.

Neural network scheme is used to pretend the human brain construction. The mind can be able to handle very complex cognitive work because of large number of neurons. Neural network has the skill of similar handling and it is composed of a group of neurons. Many researchers have proposed model on diabetes study that analyze the physiological restrictions such as body glucose attention, gender, jawbone density, power and fat tissue content , blood pressure etc. to control new healthinfo .

In this presentation Aproiri procedure has used to amount the computer aided arrangement of Breast cancer. Aproiri uses a layer through layer examine iterative process and produces the candidate groups as minor as potential. Researcher selected five breast disease data sets for rule excavating and fixed the min-support verge as 40 % and min-confidence as 80% [27-28] and then following the improved apriori method candidate itemsets were generated and rules were found that could help doctors for diagnosis of breast disease, also mined the decision instructions from breast disease analysis database and the association among breast disease and the issues such as location, oldness, harden was originate.

IV. COMPARATIVE STUDY OF APPLICATIONS

ARM techniques have become dynamic need in many existent world applications for example in Market basket analysis (MBA), commercial, medicinal field, monitoring of the patients routines etc. We have described the different applications of ARM in real life and described that how the ARM is helpful for the decision making in any field. Different Algorithms have been used by the researcher to generate the frequent patterns(FP). In MBA(market basket analysis) to describe the association between the items that a customer purchased frequent patterns(FP) are generated using APRIORI method because in APROIRI method it is easy to generates the candidates. To build the CAD of Breast cancer and to identify the most occurring diseases in a particular region Aproiri is used and it can be more efficient and fast by using F-P progress algorithm because in FP growth there is no need to scan the record every time to make the common candidates ,only 2 times scan of database is required. In Network Log data to acquire the information about the directional performance of the user SM TREE method and PD TREE method is used.

V. CONCLUSION

We proposed the practical insight about applications of association rule mining in various areas of day to day life. On one hand it provides the backend mechanism of data mining process while on the other hand it shows the output of association rule towards the day to day life, therefore it constructs a bridge between computer science and human life. The study shows that it is not only valuable for business supervision but also helpful in medical and IT industry. This paper will convey strong source for implementation and use of association rule mining ARM to comfort the process of finding trends in large data sets and will be helpful in effective and efficient decision making. The main purpose is to find the frequent patterns (FP), association, and relationship between various databases using different methods. For example in market basket analysis(MBA) ,by utilizing the association rules(AR) that are generated as a result of analyses, the retail store manager will be able to expand and apply effective marketing strategies and in disease identification frequent patterns are generated to discover the frequently occur diseases in a definite area . The conclusion in all applications is some kind of association rules (AR) that are useful for efficient decision making.

REFERENCES

- [1] Tan, P. Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. Boston: Pearson Education.
- [2] Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and Techniques* (3rd ed.). San Francisco: Morgan Kaufmann Inc.
- [3] Liao, S., & Chen, Y. (2004). Mining customer knowledge for electronic Catalog marketing. *Expert Systems with Applications*, 27, 521-532.
- [4] Chen, M., & Lin, C. (2007). A data mining approach to product assortment and shelf space allocation. *Expert Systems with Applications*, 32, 976-986.
- [5] Tan, P., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. Boston: Pearson Education
- [6] Tang, K., Chen, Y., & Hu, H. (2008). Context-based market basket analysis in a multiple-store environment. *Decision Support Systems*, 45, 150-163.
- [7] Agrawal, R., Imielinski, T., & Swami, A. (1993, May). Mining association rules between sets of items in large databases. ACM SIGMOD Conference, Washington DC, USA.
- [8] Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd Ed.). San Francisco: Morgan Kaufmann Inc.
- [9] Hao Wang, "Wireless Sensor Networks for an Extended City Intelligent Transportation System", IJACT, Vol. 3, No. 5, pp. 300 ~ 307, 2011.
- [10] He Yueshun , Du Ping, "The Research of Landslide Monitring and Pre- Warning Based on Association Rules Mining", JCIT, Vol. 6, No. 9, pp. 89 ~ 95, 2011.
- [11] Yan Hai, Xiu-li Li, "A General Temporal Association Rule Frequent Itemsets Mining Algorithm", IJACT, Vol. 3, No. 11, pp. 63 ~ 71, 2011.
- [12] B. Berendt, A. Hotho, and G. Stumme, "Toward Web Semantics", Springer LNCS 2342, pp. 264–278, 2002.
- [13] N., D. Katsaros and Y. Manolopoulos, "A data mining algorithm for generalized web prefetching," IEEE Transactions on Knowledge and Data Engineering, Vol. 15, No. 5, pp. 1155-1169, 2003
- [14] S. Chakrabarti, "Data mining for hypertext: A tutorial survey." *SIGKDD: SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery and Data Mining*, ACM, Vol. 1, No. 2, pp. 1-1, 2000
- [15] M. Balabanovic and Y. Shoham, "Learning information retrieval agents: Experiments with automated web browsing," in *Proceedings of the AAAI Spring Symposium on Information Gathering from Heterogenous Distributed Resources*, 1995, pp. 13-18
- [16] Kosala and Blockeel, "Web mining research: A survey," *SIGKDD: SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) On Knowledge Discovery and Data Mining*, ACM, Vol. 2, 2000
- [17] S. K. Madria, S. S. Bhowmick, W. K. Ng, and E.-P. Lim, "Research issues in web data mining," in *Data Warehousing and Knowledge Discovery*, 1999, pp. 303-312

- [18] J. Borges and M. Levene, "Data mining of user navigation patterns," in WEBKDD, 1999, pp. 92-111
- [19] R. Ivánçsy and I. Vajk, "Time- and Memory-Efficient Frequent Itemset Discovering Algorithm for Association Rule Mining *International Journal of Computer Applications in Technology, Special Issue on Data Mining Applications (in press)*
- [20] R. Ivánçsy and I. Vajk, "Efficient Sequential Pattern Mining Algorithms."WSEAS Transactions on Computers, Vol. 4, Num. 2, 2005, pp. 96-101
- [21] R. Ivánçsy and I. Vajk, "PD-Tree: A New Approach to Subtree Discovery.", WSEAS Transactions on Information Science and Applications, Vol. 2, Num. 11, 2005, pp. 1772-1779
- [22] J. Y Wang, H. Y Wang and D. W Zhang, et al, "Research on Frequent Itemsets Mining Algorithm based on Relational Database", Journal of Software, vol. 8, no. 8, pp. 1843-1850, 2013.
- [23] J. Y Li, J. P Wang and H. X Pei, "Data Cleaning of Medical Data for Knowledge Mining", Journal of Networks, vol. 8, no. 11, pp. 2663-2670, 2013.
- [24] X. F Chen, F C and J Wang, "Establishment of an analysis system based on clinical database of malignant blooddiseases", Journal of The Fourth Military Medical University, vol. 26, no. 21, pp. 1960-1962, 2005